# PROJECT PROGRESS REPORT
## ON
## NATURE INSPIRED OPTIMIZATION OF VIDEO POPULARITY PREDICTION

Ashish Dagar(00696303116 ), Manan Arya(01996303116), Ankit Kumar(40396303116), Saransh Taneja(40696303116)

Submitted To:- Ms. Neeti Sangwan                                    15th March 2020

## Introduction

Online Content like articles,images ,video etc. is piling or increasing each day hence our aim is to optimize the use of statistical mechanisms(regression) which are used to predict the popularity of the content. The basic idea includes optimizing feature selection using nature inspired algorithms and various other techniques; comparing all the techniques would help us reach the best possible solution.

Retrieving data from videos is computationally very expensive and time consuming. Thus, the data available at any point of time is quite minimal and feature selection using usual algorithms is tough. The motivation behind using nature inspired algorithms is to Improve the feature selection process and the success of nature inspired algorithms over the years is a testimony to the popularity of these algorithms.

Optimizing an algorithm can help in avoiding trash or unwanted features and thus ensure that the resulting model is trained on best possible features without utilizing excessive computational power for training or evaluation of the system.

Our work is divided into following phases: the initial phase comprises of training the model on a dataset comprising of social and visual features of data using support vector regression technique and the next phase involves the feature simplification using nature inspired algorithms and finally comparison of all the results lead us to the best solution for the video based prediction problem.

The aim is to study and utilize the intelligence of nature to solve the complexity regarding popularity prediction of  vast content that has been on an exponential rise in today's digitally sound world.

# PROGRESS TILL 15 MARCH

1. Upto 18th January

   Exclusive and extensive study of related papers in detail:-

   In this time period, we read a number of research papers related to our project topic and got hold of different concepts

   - Tomasz Trzcinski and Przemysław Rokita in the work titled "Predicting Popularity of Online Videos Using Support Vector Regression"
   - Sandra Paterlini and Tommaso Minnerva in the work titled "Regression Model Selection Using Genetic Algorithms"
   - Gora P., Kurach K. in work titled "Approximating Traffic Simulation using Neural Networks and its Application in Traffic Optimization."
   - Arnab Kumar Mishra, Pinki Roy, Sivaji Bandyopadhyay in the work titled "Binary particle based swarm optimization based feature selection for improving breast cancer prediction".

2. 18 January - 20 January

   Synopsis creation and submission:-

   In this time period, we created synopsis of our project and submitted to the concerned faculty.

3. 21 January - 28 January

   Collecting and Pre-Processing datasets

   In this time period, our task was to find a good dataset of videos which was created primarily for popularity prediction purposes. Through our study we got to know about the contact of a person possessing the dataset and requested for access.

   The training set consists of 1800 videos having social features of 167 days and corresponding visual features the resulting dependent variable or result is the number of views on 168th day which is an indication of popularity of the content.

4. 29 January - 15 February

   Performing popularity prediction using SVR and best parameter selection

   In this time period, we implemented the popularity of SVR over radial basis function kernels using best parameters obtained from grid search. The score obtained is saved for comparison with optimized results.

5. 16 February - 27 February

   Implementing genetic algorithm and preparing two generations for the same

   Since generations require time to pre-process we were able to run 2 generations in this time , meanwhile during the compilation of the generations we started exploring other exciting nature inspired models like crow search and have started further work on the task to validate it.

6. 7 March - 15 March

   Obtained best and average results after executing 30 generations GA

   After executing 30 generations of genetic algorithm we obtained an accuracy of approximately 85 percent which is way better than the initial accuracy of 72 percent obtained by executing SVR.

## PROGRESS INTENDED IN COMING WEEKS

1. Try to find out techniques more efficient than Genetic algorithm.
2. Implement other optimization techniques including Particle Swarm Based Optimization ,Crow Search technique and Flower Pollination Technique.